RECEIVED
CENTRAL FAX CENTER
SEP 29 2006

REMARKS/ARGUMENTS

Claims 1, 2, 9, 10, 17, 18 and 19 were amended. Claims 6-8 and 14-16 were previously withdrawn, as being drawn to a non-elected invention. The informality in claim 1, line 17 was removed.

Claims 1, 2, 9, 10, 17, 18 and 19 were amended to incorporate the organization of the data structures and dictionaries embodied in the present invention, as shown in FIG. 1, FIG. 3 and described in paragraphs [0021], [0022], [0023]- [0053] and [0062] of the present application. According to these references, each subject area 300 comprises a set of sub-subject areas 302, each sub-subject area (or subject area) comprises a set of program modules 310, each program module 310 comprises a set of arguments 320 and each argument 320 comprises a set of values 330. In the example of paragraphs [0026]-[0053] the subject area 300 is "automobiles", the program modules 310 are "sales, service and financing", the arguments 320 are "display price, submit offer, check status, and obtain financing", and the argument values 330 are "purchaser id, vehicle id, offer/price". This data organization is implemented in the dictionary, subdictionary systems of FIG. 1 (i.e., subject area dictionary, program module dictionary, argument dictionary, and value dictionary) and is mapped onto the parsed information of an unrestricted free and continuous speech natural language utterance and then used to sequentially identify parameters including subject area identifier, program module identifier, argument identifier and value identifier. These parameters are used by the computer system in order to produce computer instructions. This data structure and the process of identifying the above mentioned parameters are unique to this invention and are essential to the claimed method and apparatus for providing computer understanding and instructions from unrestricted natural language.

Claims 1, 2, 9, 10, 17, 18 and 19 were also amended to clarify that the natural language utterance is " a free continuous speech natural language" utterance, as described in the

PAGE 14/27 * RCVD AT 9/29/2006 1:04:47 PM [Eastern Daylight Time] * SVR:USPTO-EFXRF-5/2 * DNIS:2738300 * CSID:7812354409 * DURATION (mm-ss):04-22

related provisional application 60/274,786, segment P1, page 2, lines 14-15 (see Appendix A).

The Examiner rejected independent claims1, 9, 17, 18 and 19 under 35 USC 102(e) as being anticipated by Block (6,073,102). The Examiner argued that Block teaches the steps of determining a subject area identifier, determining a program module identifier, determining an argument identifier and determining a value identifier in column 8, lines-15-25, column 8, lines 30-38 and column 10, lines 30-55, column 10, 41-55, and column 15 lines 36-40, respectively.

We respectfully disagree with the Examiner's interpretation of the cited text references in Block. Referring to column 8, lines 5-23, lines 57- 62 and FIG. 3, Block teaches determining an action indicator in step 303 by employing a parser and defining an action in step 304. The definition of an action is done by "allocating a prescribable plurality of key concepts each of which respectively characterizes the action, to each action and in comparing the action indicators determined from the action information that are defined by the parameter parser PP to the key concepts. The comparison can be undertaken by a direct word comparison or, on the other hand, by an arbitrary pattern comparison." In other words, Block does not use a "context sensitive subject area dictionary system" as is the case in the present invention as claimed in claims 1, 9, 17, 18 and 19. Actually, the word "dictionary" is not mentioned anywhere in the entire Block patent. Furthermore, the term "context sensitive" has a particular meaning in the field of linguistics and computer science. According to Webster's New Millennium Dictionary, context-sensitive is defined as "in linguistics or computer syntax, pertaining to an element whose value depends on the context in which it appears." (see Appendix B). Block does not make any reference to a dictionary or to a context sensitive dictionary. Furthermore, as described in the related provisional application 60/274,786, segment P1/3 pages 2-4, " a direct" method of word recognition by "dragging" each word along a dictionary is a slow and inefficient process for developing computer understanding (see Appendix C). The present invention overcomes the inefficiencies of the direct word comparison prior art methods by structuring the data in subject areas, sub-subject areas, program modules,

arguments and values and by utilizing context-sensitive dictionaries for each subject area, sub-subject area, program module, argument and value, respectively for developing understanding of a free and continuous speech natural language utterance by a computer.

The Examiner compared the "subject areas" of the present invention to the key concepts of the Block patent, the "program modules" of the present invention to the actions, the "arguments" to the parameters and the "values" to the parameter related information. We respectfully disagree with this comparison.

Referring to column 8, lines 30-38, column 10, lines 39-55, column 15, lines 38-39, Block teaches

" A first set of <u>actions</u> is identified wherein all action indicators coincide with at least a portion of the key concepts".

" The individual actions can, for example, be characterized by the following <u>parameters</u>: Rail information (point of departure, destination, date, time of day)....."

" I would like to travel by train from Munich to Hamburg on <u>Jan.1, 1996 at 5:00</u>."

In other words, Block uses actions (characterized by key concepts), parameters, and parameter related information, i.e., three logic<u>al elements</u>. The key concepts characterize the actions and do not constitute a separate element (see column 8, lines 17-18 and column 15, lines 13-15). On the contrary, the present invention utilizes four <u>logical elements</u>, i.e., subject areas, program modules, arguments and values. Accordingly, it is not logically possible to make a direct comparison between the three logical elements of the Block patent with the four logical elements of the present invention. In order to overcome this logical discrepancy the Examiner interpreted the key concepts as a fourth element and compared it to the subject areas. We respectfully disagree with this interpretation .

Furthermore, the present invention applies to "free and continuous speech natural language utterances", whereas the Block patent refers to " a command" (column 2, lines

64-65) and "dialog arrangements" (column 1, lines 5-6). The problem with these prior art language forms is that they rely on limited number of keywords describing the key concepts for developing computer understanding. For example, referring to Block, Column 15 lines 12-32, the key concepts for action {2}, i.e., air information, are described with keywords " Fly, Flight information, Airplane". The problems with the limited number of keywords describing the key concepts for the air information action are:

a) they are limited. A speaker may use other words for air transportation ,i.e., plane, helicopter, jet, etc., which are not included in the limited set of keywords describing the key concepts

b) polisemy. Different meanings of similarly sounding words lead to confusing statements.

c) they restrict the "free and continuous speech natural language utterances"

The present invention addresses these limitations of the prior art solutions by utilizing context-sensitive dictionaries and by structuring the data in subject areas, sub-subject areas, program modules, arguments and values for developing understanding of unrestricted "free and continuous speech natural language utterances" by a computer.

In summary, the differences between the present invention as claimed in claims 1, 9, 17, 18 and 19 and the Block patent include:

- Use of context sensitive dictionaries
- Organization of dictionary data in subject areas, sub-subject areas, program modules, arguments and values.
- Direct comparison of the parsed natural language utterance with the organized context sensitive dictionaries in a stepwise process for identifying a subject area identifier, a module identifier, an argument identifier and a value identifier.
- Production of computer instructions based on four logical elements, i.e., the subject area identifier, the module identifier, the argument identifier and the value identifier.

- Production of computer instructions from unrestricted "free and continuous speech natural language utterances".

Based on these differences we conclude that claims 1, 9, 17, 18 and 19 are patentably distinguishable form the Block patent, and they should be allowed. Claims 2-5 depend upon claim 1 and claims 10-13 depend upon claim 9. Since claims 1 and 9 are patentably distinguishable from Block they should also be patcnatbly distinguishable from Block and should be also allowed.

In view of the above, it is submitted that claims 1-5, 9-13, 17, 18, 19 are in condition for allowance. Reconsideration of the rejection is requested and allowance of these claims at an early date is solicited

If this response is found to be incomplete, or if a telephone conference would otherwise be helpful, please call the undersigned at 781-235-4407

Respectfully submitted,

Aliki K. Collins, Ph.D.
Reg. No. 43,558

AKC Patents, 215 Grove Street, Newton, MA 02466
TEL: 781-235-4407, FAX: 781-235-4409

Certificate of Mailing

Date of Deposit 9/29/2006

Name: Aliki K. Collins, Ph.D. Signature

RECEIVED
CENTRAL FAX CENTER
SEP 29 2006

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as " First Class Mail" in an envelope with sufficient postage on the date indicated above and is addressed to the Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450

RECEIVED
CENTRAL FAX CENTER
SEP 29 2006

Appendix A

Provisional Patent Application

for

Method of Speaker Independent Computer Understanding of Usual Continuous Oral Speech with Very High Reliability

Inventor:

Vitaliy S. Fain

Dr.of Sci., Ph.D.

73 Outlook Dr., Apt.23
Worcester, MA 01602
U.S.A.

Tel.: (508) 797-9613
E-mail: VSFain@aol.com

Provisional application for the Main Subpatent 1 (P1)

A METHOD OF COMPUTER UNDERSTANDING USER'S NATURAL LANGUAGE INSTRUCTIONS IN A DIALOGUE PROCESS OF PROBLEM SOLVING IN AN ARBITRARY WIDE STATIONARY SUBJECT AREA

Inventor: **Professor Vitaliy S. Fain, Dr.of Sci., Ph.D.**

CLAIM:

A method of computer understanding user's natural language instructions in a dialogue process of problem solving in an arbitrary wide stationary subject area,

said method differing in that:

- a computer is supplied beforehand with the set of all program modules having participated in all programs having been used in the subject area (the stationarity of it securing such a possibility);
- a whole stationary subject area is structurized by dividing it into several subject subareas, which can be in their turn divided into sub-subareas and so on, and each sub-...-subarea is provided with a "sub-...-subarea identifier" SSI; some program modules are used in a single sub-...-subarea, the others can participate in more than one sub-subareas;
- each of the modules is provided with a "module identifier" MI, and each of the arguments of the module is provided with the "argument identifier" AI and with the "measure of argument value identifier" VI;
- the Natural Language Instruction Understanding is determined as the formation by the computer of correct i.e.expected by the user reaction to his natural language address, which reduces "understanding" just to correct recognizing the identifiers SSI, MI, AI, and VI in that address;
- structurizing the subject area leads to not too large numbers of subareas in it, sub-subareas in subareas, modules in any sub-subarea, arguments in any module, which makes the mentioned recognizing problems reasonably easy.

RECEIVED
CENTRAL FAX CENTER

SEP 2 9 2006

FUNDAMENTALS.

There are computer systems of different destinations that could be controlled, fully or partially, by orally pronounced natural language (NL) words and sentences. A subsystem of recognizing and understanding users' oral addresses - a speech understanding engine - must be a part of such a system.

This invention relates to the variety of such systems that is usually called **dialogue** systems.

We consider as dialogue such systems for computerized solving of practical (technical, medical, legal, managerial, commercial etc.) problems in which a **dialogue** is used between a user and a computer consisting of several exchanges of addresses.

In this patent application, a "dialogue" is defined as an activity of two participants when one participant passes some information to the second participant, and the latter **reacts** to it somehow, and after that the sides swop their roles.

In more detail, this invention deals with dialogues of the following kind:

Information passed to a computer by a human can be in the form of a **usual free continuous oral speech** in the user's natural language (NL). It can be referred to as the human user's instruction or address (H-address, for short) to the computer.

Still in more detail, this invention deals with just those phases of a dialogue when the human **user** gives NL instruction to the computer, and the **computer** reacts to that instruction by solving the current **subproblem** pointed out or defined by that instruction.

It is an essential property of any dialogue that a computer acts **absolutely autonomously** between any two successive human user's NL instructions called "H-addresses". An episode of autonomous work of a computer between the end of a previous H-address and the beginning of the next H-address is called "Reaction and C-address forming", C-address meaning information directed by the computer to the user.

Now, any computer can perform a work only if it has a program for it. If the previous H-address **does not contain** the description of the computer program for preparing the forthcoming reaction and C-address (which is usually the case with the dialogue systems in general and is even inevitable if H-addresses are NL), then it means that the programs of all reactions with C-addresses which can become necessary at any stage of a dialogue must be present in the computer in a **ready form**, having been put there beforehand in the form of *M* **fixed program modules**.

We call a program module "fixed" if the following items in it remain unchanged:

a) a body of a program forming it,

b) formal names or identifiers AI of arguments of a function implemented by that program

Appendix B

View results from: Dictionary | Thesaurus | Encyclopedia | the Web

Webster's New Millennium™ Dictionary of English - *Cite This Source*

Main Entry: context-sensitive
Part of Speech: *adjective*
Definition: in linguistic or computer syntax, pertaining to an element whose value depends on the context in which it appears
Example: The program offers context sensitive help.

Perform a new search, or try your search for "context-sensitive" at:

- Amazon.com - Shop for books, music and more
- HighBeam Research - 32 million documents from leading publications
- Merriam-Webster - Search for definitions
- Reference.com - Encyclopedia Search
- Reference.com - Web Search powered by Google
- Thesaurus.com - Search for synonyms and antonyms

Indicates premium content, which is available only to subscribers.

Appendix C

Provisional application for the Dependent Subpatent 1/3 (P1/3)

METHOD OF HIGHLY ACCURATE RECOGNITION OF WORDS OF CONTINUOUS ORAL SPEECH IN A MAN-TO-COMPUTER DIALOGUE SYSTEM USING RECEIVED WORDS' PROCESSING BY INDIVIDUAL ALGORITHMS FOR EACH DICTIONARY WORD AND DECOMPOSITION OF LARGE SYSTEM DICTIONARY.

Inventor: **Professor Vitaliy S. Fain, Dr.of sci., Ph.D.**

CLAIM:

A method of highly accurate recognition of words of continuous oral speech in a man-to-computer dialogue system using received words' processing by individual algorithms for each dictionary word and decomposition of the large system dictionary

differing by that:

- The method is based on extending our highly accurate word recognition Optimal Inverse Method applicable to small dictionaries to the cases of large dictionaries;
- The extension is provided by decomposing the large dictionary of a dialogue system into the structurized set of small subdictionaries, that structure reflecting the structure of the subject areas and activities in them the system deals with;
- An Optimal Inverse Method is proposed for the highly accurate recognition of words of a small subdictionary, that method being based on forming for each subdictionary word an individual word processing algorithm that takes into account the individual properties of that subdictionary word as fully as possible and is applied to a received word prior to comparing it with that subdictionary word;
- A deterministic (not full search based) method is proposed of detecting separate words in continuous speech thus making the Optimal Inverse Method applicable to the continuous speech;
- Methods of the use of mentioned large dictionary decomposition are proposed for overcoming the problems of homophony, homonymy and synonimy in computer speech understanding.

FUNDAMENTALS.

The present state of the art.

When offering any new method, it is necessary to show why the "old" ones are not good enough. In order to do so let us consider the existing methods of computer word recognition. They will be called "direct" below to stress their difference from the offered here method called "inverce".

The present invention deals with natural language dialog systems with **large vocabularies** (a large vocabulary is one containing thousands to dozens of thousand words or more).

It is known that practically all modern word recognition engines intended for continuous speech and large dictionaries are based on a Speech Units principle. The principle suggests seeing in a continuous speech signal, a sequence of "speech units", mainly speech sounds (phonemes), diphones, sillables etc. It is an attractive idea because the number of classes in the primary pattern recognition problem which is the number of speech units to be recognized is comparatively very small, usually no more than several dozens. As a result of solving that primary problem, a word initially represented by a continuous speech signal turns out to be described in terms of speech units' names, that is in a symbolic form (SF). Prototypes of words in the engine dictionary ("dictionary words") also are represented in SF. The problem of final word recognition is solved by comparing SF of a received word with SFs of dictionary words and looking for the best likeness.

Fig.1/3-1 represents a typical flow chart of oral speech word recognition system based on the "direct" method.

A speech signal representing a pronounced word comes from a microphone (1) and is subjected to some preliminary processing, for example, segmentation and initial evaluation of segments' parameters (2). The segments then undergo the main processing by means of the chosen algorithm (3) resulting in the sequence of speech units supposedly forming SF of a received word (*"received word SF"* below). This SF is compared in (4) with SF of the prototype of each dictionary word (*"dictionary word SF"* below). Presenting those dictionary word SFs for comparison in due moments and order is controlled by the unit (5) summoning them one by one from the system dictionary (6). It can be said in this case that the unit (5) performs *screening of the dictionary*, and that *each received word is "dragged" along the dictionary*.

Now it is possible to point out what is the problem with these "direct" methods. The problem consists in the inherent property of the Speech Units principle, to **entail the loss of an important for word recognition part of information** contained in a received word.

The mentioned loss of information is connected with the dual nature of an oral word.

On the one hand, different sounds in a word really form its different and distinguishable parts. They have very different physical nature, are formed in different parts of the voice tract by different sound sources and by different laws of physics. This side of the word's nature provides the understandable motivation for the Speech Parts principle and individual properties of each sound (*"local features"* below) are usually used in both forming and recognizing speech units as their distinctive features.

On the other hand, the continuity of a speech signal caused by the continuity and inertia of voice tract organs' movements leads to the fact that the properties of a sound in a word strongly depend not only on the sound itself but also on the adjacent and often even more remote sounds of a word. This side of the word's nature suggests regarding a word as something whole and therefore having the properties of a whole or integral properties. These integral properties can form strong distinctive features of a word that can be used in word recognition (*"integral features"* below). If, for example, a word contains more than one vowel, then the correlations between energies of those vowels and between their lengthes can become strong integral features of the word. There can be lots of correlations of such kind in a word, and not only between just speech units in it, but also between groups of speech units or parts of the word. It means that integral features of this kind do not belong to any one specific speech unit in a word. Both the set of those integral features and the selection of programs to compute them (this selection realizing an *"Integral Features Computing algorithm"*, IFC algorithm) are related to the *word as a whole* and are *unic* for just that word.

Obviously, the best recognition performance of an engine can be achieved if *both* the local and the integral features of words are used. It can be seen, however, that the "direct" methods do not present such a possibility.

The reason is simple. It is possible, of course, at least theoretically, to include into a dictionary word SF both its local features and its integral features. However, it is *impossible* to include the integral features into a *received word SF* and that leaves the last with only local features. Therefore, in the "direct" method, **the integral features of a word cannot participate in the comparizon of SFs** in (4, Fig.1/3-1) which results in the inevitable decrease of the quality of the recognition engine performance.

The reason of the impossibility of inclusion of integral features into a received word SF is as simple. Any attempt to compare a received word with one of the dictionary words automatically means that the hypotesis is made that the received word presents a version of the dictionary one. Therefore they must be described by the similar sets of parameters to be able to be compared, and the structure of the parameters' set is determined, at least for integral features, by the IFC algorithm of the dictionary word. However each dictionary word has *its own unic IFC algorithm*, and that means that prior to comparing with each dictionary word, the received word must be processed by the IFC algorithm of *just that* dictionary word.

So in case of a system dictionary containing, for example, several dozens of thousand words, the recognition of a *single* received word would require applying to it several dozens of thousand IFC algorithms pertaining to dictionary words. Even though it can be possible theoretically it is hardly possible practically, and in any case the "direct" methods do not go for it. Instead, *a single word SF containing only local features* is formed in (3, Fig.1/3-1) for a received word and this

received word SF in the absolutely unchanged form participates in the comparizons with all dictionary word SFs. *All the integral features remain unused.*

That is why the **existing "direct" methods** of word recognition **are not good enough indeed** and why they must be considered **non-optimal.**

Unlike this, the *"Inverse" method* of word recognition described in this Invention *does not loose any information* about a word: it uses *both its local and integral features.*

The essence of the "Inverse" method.

An oral speech address of a human user to a computer system during their dialogue will be referred to as an *H-address* below, following the denotation introduced in the Draft of the application for the Main Patent P1 ("Draft P1" below). In a dialogue, just an H-address is the source of mentioned above "received" words.

In the "inverse" method, the dictionary entry of each dictionary word includes not only its description in the form of values of its local and integral features but also references to program modules realizing the IFC algorithm of that word (*"IFC-program"* below). When the moment comes to compair a received word with the prototype of some dictionary word, not only local features of the received word are calculated but it is subjected to the processing by the IFC-program attached to that dictionary word, this processing creating the integral features of the received word in the terms of that dictionary word. After that both local and integral sets of features of the two words are compared and the degree of their resemblance calculated.

After that is accomplished, any one of the two possible modes of further operating can be pursued:

a) *The same received word* from the H-address passes to the next dictionary word and confronts with it. This is just the mode used in the "direct" methods and called "screening the dictionary " above except that in this case the received word each time undergoes a new processing by the IFC-program of *each* next dictionary word.

b) *The same dictionary word* passes to the next received word from the H-address and confronts with it.

Of course, both these modes preserve the integral features and the possibility to use them in the word recognition. The word **"Optimal"** in the name of the proposed in this Invention word recognition method (as opposed to the "non-optimality" of "direct" methods) reflects this fact. As for the usefuness of each of the two modes, just **the mode b) is suggested in this invention.**

The reason of this choice is quite simple. Let the number of dictionary words be N, the number of received words in an H-address be W.

In the mode a) the number of necessary reloadings of IFC-programs is obviously NW. At the same time, it is as obvious that in the mode b) that number is only N.

It can be seen that the way the received words confront with the dictionary words in the mode b) is in a sense *opposite* to the one in the mode a). That is why the proposed in this Invention word recognition method is called **"Inverse"**. It seems also natural to call the process of confronting